WOODHOUSE EXHIBIT 9

UNREDACTED VERSION OF JOINT LETTER BRIEF [DKT. 357]

FILED UNDER SEAL

December 20, 2024

E-Filed

The Honorable Thomas S. Hixson
United States District Court for the Northern District of California
San Francisco Courthouse, Courtroom E – 15th Floor
450 Golden Gate Avenue
San Francisco, CA 94102

Re: *Kadrey et al. v. Meta Platforms, Inc.*; Case No. 3:23-cv-03417-VC-TSH

Dear Judge Hixson:

      The parties jointly submit this letter brief regarding issues related to Meta's Supplemental Responses to Plaintiffs' First, Second, and Third Sets of Interrogatories served on December 13, 2024. The parties met and conferred on December 17 but were unable to reach a resolution.

I.   PLAINTIFFS' POSITION

Late on December 13, 2024, Meta served supplemental responses to all three sets of Plaintiffs' interrogatories. Some of those responses are deficient. Others create new issues that warrant further amendment. Still others ignore this Court's prior orders on relevance and construction of defined terms. *See* Dkt. 315. In light of these issues, Plaintiffs request the following relief.

A. Supplemental Responses to First Set of ROGs, Ex. A at 5-46.

- Llamas 4&5: Meta's supplemental ROG responses served 12/13 are still limited to Llamas 1-3, even though this Court already held that Llamas 4 and 5 are also relevant. Dkt. No. 279 at 4 ("Llama 4 is relevant to this case, notwithstanding that it is still under development."); Dkt. No. 315 at 7 ("Llama 5 is relevant"). It should not matter that the ROGs initially defined only Llamas 1-3: they were issued before Plaintiffs had any knowledge of the existence of Llamas 4-5. The entire point of Rule 26(e)'s duty of supplementation is to ensure discovery responses "reflect new information" that developed afterwards. *Abernathy v. Liberty Ins. Corp.*, 2023 WL 6785797, at *2 (C.D. Cal. Sept. 22, 2023). In the intervening time period, Meta has been rapidly developing Llamas 4-5, and the Court held they're relevant. This constitutes "new information," and a proper supplementation must update the ROG answers for the new models.

- Agreements: Meta continues to improperly limit the term "Agreements" to written contracts, despite this Court's rejection of that limitation and adoption of Plaintiffs' broader definition. Dkt. 315 at 8-9.

- Training Data: Meta improperly limits the definition of "Training Data" to Books3. But there are many other datasets that Meta uses as training data in addition to Books3, as well as data obtained from web crawling. Meta must expand its responses. At the time Plaintiffs served these ROGs (before any document productions), Books3 was the only shadow/pirated dataset that Meta had disclosed as constituting part of its training data. *See* Dkt. 69. Since discovery has revealed a treasure trove of "new information" about Meta's use of other pirated datasets and works as training data, there is no basis to continue limiting the ROG responses to Books3.

- ROG 1: Meta's response produces a chart of datasets along with narrative explanation. Yet neither the chart nor the explanation includes Llamas 4-5 despite their court-ordered relevance.

- ROG 2: Meta did not include this in their 12/13 ROG responses. Since Meta did reproduce other ROG responses that were not supplemented on December 13, Meta must confirm that its response to ROG 2 served on August 22, 2024 continues to be the final response to this ROG.

- ROGs 3 & 7: Meta's responses do not appear to include Llamas 4 and 5. Regardless, the responses fail to provide details about how *any* Llama model was/is fine-tuned. The responses similarly are scant on details of the specifics of Meta's efforts around memorization for any model.

- ROG 4: Meta's response fails to identify/explain all the "risks" it references or "mitigation(s)" it employed.

- ROG 8: Meta's response fails to address any of Llamas 3, 4, and 5. It must address all of them.

- ROG 10: Meta improperly limits the definition of "Training Data" to Books3. Further, this ROG asks Meta to identify and supply information about all Persons—which, as defined, includes hosts of online pirated datasets, *see* Ex. A at 4—from whom it obtained training data.

1

Yet Meta fails to provide all the requested information for *any* datasets it copied, let alone all of them. Meta must do so, including identifying dates it obtained each copy of pirated datasets.

- ROG 13: Meta's response does not address Llamas 4 or 5. Further, mere reference to Bates numbers does not answer the entire ROG. In addition to identification of all training datasets, the ROG also requests a list of any forbidden datasets and reasons why they were prohibited.

B. Supplemental Responses to Second Set of ROGs, Ex. A at 47-66.

- Llamas 4&5: Meta's objections continue to improperly limit Meta's responses to Llamas 1-3.

- Agreements: Meta still improperly limits the definition of "Agreements" to written contracts.

- Training Data: Meta again improperly limits the definition of "Training Data" to just Books3.

C. Supplemental Responses to Third Set of ROGs, Ex. A at 67-85.

- Llamas 4&5: Meta's objections continue to improperly limit Meta's responses to Llamas 1-3.

- Shadow Datasets: Meta hasn't disclosed all datasets it uses with Llamas 4 and 5, and thus improperly restricts the definition of "Shadow Datasets" to the extent it excludes any used with those models.

- ROG 19: Meta fails to set forth revenue it anticipates receiving with its ███████ ██████ (or even a projection). The ROG expressly asks for expected as well as actual revenue.

- ROG 22: This ROG asks whether Meta made any efforts to obtain licenses to use Shadow Datasets or the works therein. Meta responds with a colloquy that it didn't need permission. The ROG does not ask Meta for its opinion about what it needed to do or didn't need to do. If Meta didn't do anything, then it should just say so, and that's the end of it. *See* Dkt. 315 at 5.

- ROG 23: Meta's response fails to identify what datasets it obtained from Anna's Archive, "a search engine for Shadow Libraries."[1] Meta's answer is akin to citing Google in an academic paper because the sources were initially found via Google Search. It's also unclear from Meta's response if it obtained datasets from other sources for any Llama models or if it's just the three listed.

D. Plaintiffs Timely Raised These Deficiencies in Meta's Interrogatory Responses.

During the parties' M&C, Meta suggested Plaintiffs are time-barred from challenging its supplemental responses to ROG Sets 1 and 2 because those ROGs were "Existing Written Discovery." Dkt. No. 258. The parties should not revive this long-squabbled-over term "like some ghoul in a late night horror movie that repeatedly sits up in its grave and shuffles abroad, after being repeatedly killed and buried." *Lamb's Chapel v. Center Moriches Union Free Sch. Dist.*, 508 U.S. 384, 398 (1993) (Scalia, J., concurring). Regardless, this argument is nonsensical. Meta amended its responses to include new information and, as discussed above, they now conflict with this Court's recent orders. This necessarily renders Plaintiffs' challenges—brought within one week of Meta's service of its supplemental responses—timely and appropriate. Otherwise a party would have no recourse to challenge the other's supplemental discovery responses. But that is not the rule.

---

[1] https://en.wikipedia.org/wiki/Anna%27s_Archive

II.   META'S POSITION

The issues above, which pertain to discovery served long ago, were not raised in the serial discovery motions filed by Plaintiffs over the past several weeks, instead they were withheld until Dec. 16, after the close of fact discovery. Like the other letter briefs filed today, Plaintiffs' manufactured discovery disputes raise issues that are baseless, irrelevant, and/or time-barred.

First, as to Meta's responses to Plaintiffs' 1st and 2nd Sets of Interrogatories ("ROGs"), Plaintiffs have "once again…fail[ed] to meet the deadline to present disputes regarding 'Existing Written Discovery.'" Dkt. 253. As the Court knows, Plaintiffs were obligated to bring disputes on Existing Written Discovery by Oct. 23, 2024, later extended to Nov. 9, 2024 (Dkt. 253). The 1st and 2nd Sets of ROGs, to which Meta first responded on Feb. 23, 2024 and Sept. 30, 2024, respectively, were clearly "Existing Written Discovery." If Plaintiffs had concerns regarding those responses, they should have timely brought them by Nov. 9, 2024, not on the last day to bring discovery disputes. Meta has identified each of these interrogatories as "**Waived**" below.

Meta's supplementation of certain responses for the 1st and 2nd Sets of ROGs does not render Plaintiffs' motion timely. Plaintiffs' issues could have been brought months ago, but were not. For instance, Plaintiffs take issue with Meta's consistent *objections* to certain definitions to the 1st and 2nd Sets of ROGs, served in February and September. *Compare*, Ex. B (Meta's Responses to ROGs, Set One, served Feb. 23, 2024), Ex. C (Meta's Responses to ROGs, Set Two, served Sep. 30, 2024). Moreover, the details Plaintiffs now claim are missing were not included in prior responses to the 1st and 2nd Sets of ROGs, thus there were no "new" issues in those responses for Plaintiffs to raise here.

Equally puzzling is that much of Meta's supplementation of its responses to the 1st and 2nd Sets of ROGs was the result of prior meet and confers and compromises reached between the parties. *See*, *e.g.*, Ex. D (Oct. 2024 emails reflecting agreements as to ROGs 3, 4, and 13). As to Plaintiffs' 3rd Set of ROGs, the parties met and conferred on Meta's initial responses to those ROGs weeks ago, and Meta supplemented its responses to avoid motion practice. *See, e.g.*, Ex. E at 1 (Nov. 26, 2024 email confirming parties' dispute as to ROGs 19, 20 and 22 were "moot"). Other than ROG 22, none of the issues raised then included the issues Plaintiffs raise now.

While the motion should be denied as untimely, it should also be denied on the merits. Plaintiffs fail to address or even acknowledge Meta's objections to the ROGs at issue. These include, but are not limited to, objections based on undue burden, proportionality, relevance, and privilege. Plaintiffs do not challenge these objections, which are proper and appropriately limit Meta's responses. Nor do Plaintiffs explain the relevance or proportionality of the additional information they seek in response to ROGs 3, 4, 7, 10, 13, 19, or 23 (apart from information relating to Llama 4 and 5, discussed below).

As to Llama 4 and 5,[2] Plaintiffs' reliance on this Court's prior orders concerning Meta's RFP responses is unavailing. Plaintiffs' 1st and 2nd Set of ROGs define "Meta Language Models" in reference to Llama 1-3–they do not encompass Llama 4 and 5. *See* Ex. F (excerpts of Plaintiffs' 1st and 2nd ROGs). Furthermore, Meta's responses to Plaintiffs' 2nd and 3rd Sets of ROGs are not limited to Llama 1-3. Plaintiffs are also far past the deadline to bring disputes concerning the 1st

---

[2] In addition to Plaintiffs' challenges to Meta's objections as to all ROGs, Plaintiffs' challenges to Meta's responses to ROGs 1, 3, 7, 8, and 13 are also based, in whole or in part, on Meta's purported failure to address or include Llama 4 and 5.

Set of ROGs. In any event, those orders only required that Meta not exclude or decline to produce responsive *documents* that relate to Llama 4 and 5, which would have been encompassed by Meta's existing searches. Dkt. 279 at 4; Dkt. 315 at 6. They did not pertain to Meta's ROG responses (on which Plaintiffs never moved), let alone require that Meta supplement them.

Meta addresses the remaining issues raised by Plaintiffs below:

Definitions of "Agreements" and "Training Data" (**Waived**):   As Meta has repeatedly represented to Plaintiffs, Meta did not restrict its ROG responses based on its limited construction of these terms.

ROG 1 (**Waived**): ROG 1 defined "Meta Language Models" as Llama 1-3, and Meta responded accordingly. Nevertheless, Meta has produced, and in ROG 13 referred to, documents reflecting various other datasets that have been or are being considered for use with future Llama models. To the extent Meta is ordered to further amend its ROG responses, Meta refers Plaintiffs to Meta_Kadrey_00185698, which Meta produced on December 13, 2024 and includes the list of datasets decided on for use to train Llama 4.[3]

ROG 2 (**Waived**): Meta confirms that its response to ROG 2 served on *August 22, 2024* continues to be its final response to this ROG.

ROGs 3 & 7 (**Waived**): Meta made several objections to these ROGs, which Plaintiffs do not address. Contrary to Plaintiffs' demands, this ROG has nothing to do with memorization, and Meta, in fact, provided information regarding finetuning (as previously agreed upon), referring to documents and deposition testimony. As to ROG 7, again, the ROG itself defined Meta Language Model as Llama 1-3, and Meta responded accordingly. To the extent Plaintiffs attempt to rewrite the ROG to cover additional Llama models, it would be unduly burdensome and disproportionate for Meta to identify all of the individuals involved in Llama 4 and 5, which are still in development (as opposed to the released models which have published papers identifying those who were involved). Even aside from this being a moving target, Meta would need to conduct an extensive investigation to identify these individuals, and with fact discovery closed and Plaintiffs having deposed dozens of Meta witnesses, any conceivable utility of such information is marginal at best.

ROG 4 (**Waived**): Meta's supplemental response identifies the risks that Meta witnesses testified to in deposition and directs Plaintiffs to relevant deposition testimony. Meta agreed to supplement its response following a meet and confer in October, which, at the time, Plaintiffs accepted as sufficient. Most of the identified risks (data privacy, product safety, social risk, potential abuse by bad actors) have nothing to do with the issues in the case and would thus be disproportionate to the needs of the case. Moreover, plaintiffs do not identify the scope of "explanation" of risks they seek, which may invade the attorney-client privilege because the risks intersect with legal issues. Finally, this interrogatory does not request anything about "mitigations," which, in any event, were described in cited deposition testimony.

ROG 8 (**Waived**): Meta never supplemented its response to ROG 8, which was served in February. Plaintiffs never raised issues concerning this ROG, and are barred from doing so now after the deadline to bring disputes concerning "existing written discovery" has passed.

ROG 10 (**Waived**): Meta provided responsive information within its possession, custody, and control, including the sources of the datasets it obtained and licensing agreements for training data.

---

[3] Given the nascency of the model, no similar list exists for Llama 5.

4

Until now, Plaintiffs never demanded that Meta identify the "*hosts* of online pirated datasets," a vague and ambiguous term which Plaintiffs do not define, let alone explain the relevance of.

ROG 13 (**Waived**): Meta supplemented this response to the extent that it had previously agreed to, namely to identify datasets considered for use with Meta's Llama models (including future models) and indicating which were approved. To the extent Plaintiffs seek additional information, Meta made several objections, including, specifically, privilege objections.

"Shadow" Datasets: Meta has produced the list of datasets for training Llama 4 and documents showing datasets considered for use with future models. "Plaintiffs can tell Meta which of those are Shadow Datasets" (Dkt. 315). They did not.

ROG 19: Meta amended its response to "provide a high-level explanation of the identified documents and what they show regarding Meta's revenue from partnerships," as agreed-upon by the parties. *See* Ex. E at 2.

ROG 22: Plaintiffs previously agreed not to move on this ROG, and Meta has not supplemented its response to this ROG. *See id.* at 1.

ROG 23: The information Plaintiffs seek – what datasets Meta obtained from Anna's Archive and whether it obtained datasets from sources other than those provided in Meta's response – is plainly not called for by this ROG. It is also irrelevant. Meta has otherwise responded with information that could be located following a reasonable investigation.

III.   PLAINTIFFS' REPLY

Meta's "waiver" objection is meritless. To start, Meta ignores Judge Chhabria's order that "the deadline to raise disputes regarding additional discovery"—which includes these supplemental ROGS served on 12/13—remains 7 days after the close of discovery[.]" Dkt. 253. Meta's supplemental ROG responses are not "Existing Writing Discovery"; indeed, Meta ignores its Rule 26(e) duty to supplement incomplete responses, which plainly includes those that omit Llamas 4 and 5. Moreover, Meta cites no authority to support its "waiver" arguments (nor is there any).

Meta failed to supplement responses to Plaintiffs' 1st ROG Set, including as to Llamas 4&5. Meta incorrectly states that Plaintiffs' definition of Llama Models in their 1st Set of ROGs doesn't cover Llamas 4 or 5 (which are relevant). The definition includes all "variant models," precisely to capture models like Llamas 4 and 5 developed subsequent to the issuance of these ROGs.

ROGs 1&10. Meta's listing of Llama 4 datasets two hours before the close of discovery does not come close to satisfying the ROGs. Since Meta finally identified the Llama 4 books datasets, it can't dispute information about them is relevant. Meta must supplement the ROGs to provide the requested information, including identifying all copies of the books datasets and when it took them.

ROGs 4&13. Meta does not even pretend its responses provide all the information requested, instead citing its rote, boilerplate objections. Those are meritless, as are its unexplained references to privilege—but if Meta intends to invoke privilege to withhold information, it needs to be precise.

ROGs 19&22. Meta suggests Plaintiffs backtrack on a prior decision not to challenge Meta's responses. Not so. Meta promised to amend to resolve Plaintiffs' concerns. It then failed to do so.

ROG 23. Meta's response makes no sense. It expressly asks about all sources from which Meta obtained its shadow (i.e. pirated) datasets. Meta even identifies Anna's Archive, but that's just a search engine for shadow libraries, i.e. sources of shadow datasets, none of which Meta identifies.

5

By: /s/ *Bobby Ghajar*

Bobby A. Ghajar
Colette A. Ghazarian
COOLEY LLP
1333 2nd Street, Suite 400
Santa Monica, CA 90401
Telephone: (310) 883-6400
Facsimile: (310) 883-6500
Email: bghajar@cooley.com
       cghazarian@cooley.com

Mark R. Weinstein
Elizabeth L. Stameshkin
COOLEY LLP
3175 Hanover Street
Palo Alto, CA 94304
Telephone: (650) 843-5000
Facsimile: (650) 849-7400
Email: mweinstein@cooley.com
       lstameshkin@cooley.com

Kathleen R. Hartnett
Judd D. Lauter
COOLEY LLP
3 Embarcadero Center, 20th Floor
San Francisco, CA 94111
Telephone: (415) 693-2071
Facsimile: (415) 693-2222
Email: khartnett@cooley.com
       jlauter@cooley.com

Phillip Morton
COOLEY LLP
1299 Pennsylvania Avenue, NW, Suite 700
Washington, DC 20004
Telephone: (202) 842-7800
Facsimile: (202) 842-7899
Email: pmorton@cooley.com

Angela L. Dunning
**CLEARY GOTTLIEB STEEN &
HAMILTON LLP**
1841 Page Mill Road, Suite 250
Palo Alto, CA 94304
Telephone: (650) 815-4121

By: /s/ *Maxwell V. Pritt*

BOIES SCHILLER FLEXNER LLP
David Boies (*pro hac vice*)
333 Main Street
Armonk, NY 10504
(914) 749-8200
dboies@bsfllp.com

Maxwell V. Pritt (SBN 253155)
Joshua M. Stein (SBN 298856)
44 Montgomery Street, 41st Floor
San Francisco, CA 94104
(415) 293-6800
mpritt@bsfllp.com
jischiller@bsfllp.com
jstein@bsfllp.com

Jesse Panuccio (*pro hac vice*)
1401 New York Ave, NW
Washington, DC 20005
(202) 237-2727
jpanuccio@bsfllp.com

Joshua I. Schiller (SBN 330653)
David L. Simons (*pro hac vice*)
55 Hudson Yards, 20th Floor
New York, NY 10001
(914) 749-8200
dsimons@bsfllp.com

Interim Lead Counsel for Plaintiffs

6

Facsimile: (650) 849-7400
Email: adunning@cgsh.com

Attorneys for Defendant Meta Platforms, Inc.

8

ATTESTATION PURSUANT TO CIVIL LOCAL RULE 5-1(h)

I hereby attest that I obtained concurrence in the filing of this document from each of the

other signatories. I declare under penalty of perjury that the foregoing is true and correct.

Dated: December 20, 2024   BOIES SCHILLER FLEXNER LLP

                                                         /s/ Maxwell V. Pritt
                                                         Maxwell V. Pritt
                                                         Reed Forbush
                                                         Jay Schuffenhauer

                                                         Attorneys for Plaintiffs